DRES

DEFENCE RESEARCH ESTABLISHMENT SUFFIELD

SM 1481

UNCLASSIFIED

NATO Panel VII/SICA International Training Exercise on the Identification of Peptides

BY

J.R. Hancock and P.A. D'Agostino

SEPTEMBER 1996

CRAD

National Defence Défense nationale

Canada

Suffield Memorandum 1481

NATO Panel VII/SICA International Training Exercise On The Identification of Peptides

by

J.R. Hancock and P.A. D'Agostino

Executive Summary

Title: J.R. Hancock and P.A. D'Agostino, " NATO Panel VII/SICA International Training Exercise On The Identification of Peptides", Suffield Memorandum No. 1481, 1996, UNCLASSIFIED.

Introduction: The Canadian Forces (CF) may be called on to perform peacekeeping or peacemaking operations in regions of the world where there is a significant threat of chemical/biological warfare agent use. To operate effectively in these theaters the CF must be able to identify the exact nature of the chemical/biological agent(s). Mass spectrometry (MS) is a powerful analytical technique for the identification of both known and unknown compounds and DRE Suffield, in conjunction with its NATO allies, is currently investigating this instrumental technique in fulfilment of CF and NATO agent detection and identification requirements.

Results: Within the NATO defence community, the Terms of Reference for the Sampling and Identification of Chemical/Biological Agents (SICA) subgroup have recently been revised to include the entire CB agent spectrum. The lack of analytical methods for the identification of mid-spectrum agents prompted SICA to hold an international training exercise focussing on these agents. From the 11 NATO countries represented on the SICA subpanel, laboratories from Canada (host nation), Denmark, Netherlands, Norway, United Kingdom and United States (2 laboratories) agreed to participate in the international training exercise. The objective of the exercise was to evaluate the basic capabilities of the laboratories to determine molecular weights and primary amino acid sequences for five unknown peptides with molecular weights in the same mass range as many mid-spectrum agents.

All the participating laboratories demonstrated the basic ability to determine either monoisotopic or average molecular weights for the unknown peptide samples. The most accurate results, with an average error of 9.2 ppm, in this part of the training exercise were reported by Canada using electrospray ionization interfaced to a high resolution mass spectrometer. The second objective of the training exercise was to determine partial or complete amino acid sequences for the unknown peptides. The only laboratory to provide the correct amino acid sequence for all five peptides was the United States #1 laboratory. The laboratory from the Netherlands using only mass spectrometry was able to completely sequence four of the unknown peptides and provided partial sequence data for the fifth peptide.

Significance of Results: The CF may be deployed in regions of the world where there is a significant threat of chemical/biological warfare agent use. Identification of the agent is of importance since the results of such analyses would contribute to the development of strategic and political positions regarding future Canadian military operations and would facilitate the dissemination of technical advice to in-theater field commanders and medical personnel.

Future Goals: The CB threat spectrum includes chemical and biological warfare agents and toxins of biological origin in the "mid-spectrum" between these agents. The identification research effort has been focused on the detection and identification of these toxins of biological origin. Use of these warfare agents could easily go unconfirmed, as analytical methods have not been fully developed for their identification. DRE Suffield is now actively addressing this deficiency through the application and development of MS methods for the identification of these agents.

ABSTRACT

The revision of the Terms of Reference for NATO's Sampling and Identification of Chemical/Biological Agents (SICA) subgroup to include the entire CB agent spectrum and the current lack of analytical methods for mid-spectrum agents prompted SICA to hold an international training exercise focussing on the identification of these agents. From the 11 NATO countries represented on the SICA subpanel, laboratories from Canada (host nation), Denmark, Netherlands, Norway, United Kingdom and United States (2 laboratories) agreed to participate in the international training exercise. The objective of the exercise was to evaluate the basic capabilities of the laboratories to determine molecular weights and primary amino acid sequences for five unknown peptides with molecular weights in the same mass range as many mid-spectrum agents.

The determination of an unknown's molecular weight is considered to be one of the initial steps in the identification of a mid-spectrum agent. All the participating laboratories demonstrated the basic ability to determine either monoisotopic or average molecular weights for the unknown peptide samples. The most accurate results, with an average error of 9.2 ppm, in this part of the training exercise were reported by Canada, using electrospray ionization interfaced to a high resolution mass spectrometer.

The second objective of the training exercise was to determine partial or complete amino acid sequences for the unknown peptides. The only laboratory to provide the correct amino acid sequence for all five peptides was the United States #1 laboratory. They were able to sequence the unknown peptides by using a combination of an automated peptide sequencer and sequence data provided by mass spectrometry. The laboratory from the Netherlands using only mass spectrometry was able to completely sequence four of the unknown peptides and provided partial sequence data for the fifth peptide.

CONTRIBUTING LABORATORIES

CANADA

Defence Research Establishment Suffield
PO BOX 4000, Medicine Hat, Alberta, Canada, T1A 8K6

Contributors: Mr. J.R. Hancock
Dr. P.A. D'Agostino

DENMARK

Ministry of the Interior
Emergency Management Agency, Chemical Laboratory
2 Universitetsparken, DK-2100, Copenhagen, Denmark

Contributors: Mr. M. Thomsen
Mr. J.S. Jensen

NETHERLANDS

Prins Maurtis Laboratory - TNO
P.O. Box 45, 2280 AA Rijswijk, The Netherlands

Contributors: Dr. Ch. E. Kientz
Mr. A. Hulst

NORWAY

Norwegian Defence Research Establishment
P.O. Box 25, 2007 Kjeller, Norway

Contributors: Dr. B.A. Johnsen
Dr. L.H. Bjerkeseth

UNITED KINGDOM

CBDE Porton Down
Salisbury, Wiltshire, SP4 OJQ, United Kingdom

Contributors: Dr. A. Trethewey
Dr. R. Black

UNITED STATES

Laboratory #1
Research & Technology Directorate
US Army Edgewood RD&E Center, Aberdeen Proving Ground
MD 21010-5423, United States

Contributors: Dr. T. Krishnamurthy
Dr. M. Prabhakaran
Dr. P.L. Ross

Laboratory #2
Edgewood Research, Development & Engineering Center
Attn: AMSCB-ACL, Aberdeen Proving Ground
MD 21010-5423, United States

Contributors: Dr. P. Smith

INTRODUCTION

NATO may be called upon to deploy military forces in support of peacekeeping/peacemaking or battlefield operations in regions of the world where there is a significant threat of chemical/biological warfare (CBW) agent use. To operate effectively in these theatres, NATO forces must be able to detect and identify CBW agent(s). NATO doctrine states that as an alliance of 16 nations, it takes the consensus of all nations before NATO responds to the use of CBW agents against NATO troops. Consensus would only be reached if all the evidence of CBW agent use, including the identification process, clearly supported the allegation.

Under the umbrella of the NATO Army Armaments Group (NAAG), Panel VII on NBC Defence established a group of experts to deal with the problems associated with the Sampling and Identification of Chemical Agents (SICA). This group produced Allied Engineering Publication 10 (AEP-10), a series of books dealing with procedures and techniques for sample collection, packaging, transport and identification of samples believed to contain chemical warfare agents. National representatives from 11 NATO countries meet annually to review progress in achieving this broad objective and to provide the military with advice and guidance on sampling and identification techniques which can be used to provide unambiguous identification of chemical warfare agents.

In 1995, Panel VII amended SICA's terms of reference to include sampling and identification of the entire CB agent spectrum in recognition of the broader threat to NATO troops. The CB agent spectrum encompasses a wide range of toxic materials including chemical warfare agents, mid-spectrum agents (e.g., toxic peptides and proteins) and biological warfare agents. Recent advances in biotechnology, including the isolation and production of peptides using solid-phase synthesis and recombinant DNA-modified microorganisms, has opened up new avenues for the preparation of militarily significant quantities of agents in the "mid-spectrum" between classical chemical and

classical biological warfare agents. The lack of appropriate analytical methods makes detection and identification of mid-spectrum agents extremely difficult and would likely result in inconclusive evidence.

At the May 1995 meeting of SICA, the delegates discussed a proposal presented by Canada to hold a SICA international training exercise in 1996. With the expanded mandate of SICA, it was felt that there would be considerable benefit to NATO if SICA held a training exercise dealing with the identification of mid-spectrum agents. Canada agreed to host and participate in an exercise designed to evaluate the basic analytical capabilities of the NATO laboratories. Each laboratory was provided with five unknown peptides with molecular weights between 200 and 5000 daltons (Da), which encompasses the molecular weights of many mid-spectrum agents. The two objectives of this training exercise were:

a) Determine the molecular weight (monoisotopic or average) for five unknown peptides and

b) Using available techniques, determine partial or complete amino acid sequences for the unknown peptides.

Seven national laboratories from Canada, Denmark, Netherlands, Norway, United Kingdom and United States (2 laboratories) agreed to participate in the international training exercise. The samples were distributed to the laboratories by 1 February 1996 with the constraint that the analysis be completed by the end of the month. Within that timeframe, a limit of five working days was placed upon the analysis of the samples.

EXPERIMENTAL

Sample Description

The peptides used in this international training exercise were synthesized and distributed by Dr. R. Hodges' Group (Protein Engineering Network of Centres of Excellence, University of Alberta, Edmonton, Alberta, Canada) as lyophilized powders. Table I lists the information provided to each laboratory by the University of Alberta.

Table I. Information Provided by University of Alberta on SICA Unknown Peptides

Peptide	Molecular Weight Range	N-Terminal	C-Terminal
A	200-800	free	free
B	900-1400	free	free
C	1400-2100	free	free
D	2300-2900	free	amide
E	4100-4700	free	free

Sample Handling

It was suggested to the laboratories that upon receipt of the samples that they be refrigerated until the beginning of the exercise. Each sample vial contained approximately 1mg of unknown peptide and the laboratory was free to reconstitute the sample with the solvent system of their choice. The solvent systems used by the participating laboratories included, distilled water or distilled water with a small percent of acid (0.2% formic, 0.05% acetic) , methanol/water (50/50) with 0.5% acetic acid and acetonitrile/water (75/25) with 0.1% trifluoroacetic acid.

A number of laboratories carried out enzymatic digestions of the unknown peptides in order to produce lower mass fragments from which they could obtain sequence data. The most commonly used enzyme was trypsin with two laboratories also using chymotrypsin. The United States #1 laboratory also used Glu-C and methylation in order to determine the number of acidic sites.

Instrumentation

All the participating laboratories used mass spectrometry to determine the molecular weights for the unknown peptides. Table II list the mass spectrometric instrumentation and ionization modes used by the individual laboratories. Amino acid sequence information, where reported, was generally obtained by mass spectrometry, with the United States #1 laboratory also using an Applied Biosystems 477A Automated Peptide Sequencer.

Table II. Mass Spectrometric Instrumentation

Country	Ionization Mode	MS Instrumentation
Canada	Electrospray	VG Autospec Q Hybrid Tandem MS
Denmark	Electrospray	Finnigan-MAT LCQ MS
Netherlands	Electrospray	VG Quattro II MS
Norway	Electrospray	Finnigan MAT95Q Hybrid Tandem MS
United Kingdom	Electrospray FAB	Finnigan-MAT TSQ-700 MS VG Autospec Q Hybrid Tandem MS
United States #1	Electrospray MALDI FAB	Finnigan-MAT TSQ-700 MS Vestec 2000 TOF MS JOEL HX110/HX1110 Tandem MS
United States #2	Electrospray	Finnigan-MAT TSQ-7000 MS

RESULTS AND DISCUSSION

Determination of Molecular Weight

The primary objective of the SICA international training exercise was to evaluate the ability of each laboratory to determine molecular weights of unknown peptides. Table III lists the amino acid sequences and monoisotopic and average molecular weights for each of the peptides labelled A-E.

TABLE III. Amino Acid Sequences and Molecular Weights for SICA Unknown Peptides

Peptide	Amino Acid Sequence	Monoisotopic Molecular Weight (Da)	Average Molecular Weight (Da)
A	FIPK	503.310755	503.642077
B	AGKDYDKIEE	1166.545505	1167.237777
C	ATKKEVPLGVAADANKLG	1781.004645	1782.070277
D	RFEMFRELNEALELKDAQAGKE-NH_2	2622.322369	2623.966791
E	VEHYDNIEQKIDDIDHEIADLQAKITRLVQQ HPRIDE	4436.235975	4438.877777

All participating laboratories used mass spectrometry to determine either monoisotopic or average molecular weights for the unknown peptides. In some cases, such as the United Kingdom laboratory and the United States #1 laboratory, more than one mass spectrometric technique was used and their results were submitted separately. Only the Canadian laboratory provided information on the precision of the analysis and number of replicate analyses. Table IV lists the molecular weights submitted by each laboratory and the individual as well as average ppm error for the exercise.

Table IV. Monoisotopic and Average Molecular Weights Reported Participating Laboratories

Country	Peptide A Monoisotopic MW 503.310755 Average MW 503.642077		Peptide B Monoisotopic MW 1166.545505 Average MW 1167.237777		Peptide C Monoisotopic MW 1781.004645 Average MW 1782.070277		Peptide D Monoisotopic MW 2622.322369 Average MW 2623.966791		Peptide E Monoisotopic MW 4436.235975 Average MW 4438.877777		Average ppm Error (n=5)
	M.W.	Error (ppm)	M.W.	Error (ppm)	M.W.	Error (ppm)	M.W.	Error (ppm)	M.W.	Error (ppm)	
Canada[a]	503.312 ±0.001* (n=3)	2.5	1166.550 ±0.006* (n=3)	3.9	1780.991± 0.006* (n=3)	7.7	2622.276 ±0.005* (n=3)	18	4436.172 ±0.034* (n=4)	14	9.2± 6.6
Denmark[a]	503*	617	1166*	468	1782*	559	2622*	123	4437*	172	387±226
Netherlands[a]	503.4*	177	1166.6*	47	1781.2*	110	2623.8	64	4438.6	63	92±53
Norway[a]	503*	617	1167*	390	1781	601	2623	368	4439	28	401±238
United Kingdom[a]	503.4	481	1166.7	461	1781.4	376	2623.8	64	4438.2	153	307±188
United Kingdom[c]	503.1	1076	1166.6	546	1780.9	657	2623.8	64	4437.8	243	517±392
United States #1[a]	503.9	512	1167.6	310	1784.6	1420	2624.2	89	4438.4	108	488±549
United States #1[b]	503.4	481	1167.1	118	1781.3	432	2624.5	203	4437.8	243	295±155
United States #1[c]	503.62*	614	1166.64*	81	1781.33*	183	2623.10*	297	4436.41*	39	243±230
United States #2[a]	503.4*	177	1166.7*	132	1781.3*	166	2622.8*	182	4437.7	265	184±49

[a] ESI, [b] MALDI, [c] FAB
* Monoisotopic Molecular Weight

Electrospray (ESI) mass spectrometric data were reported by all laboratories with several laboratories also proving matrix assisted laser desorption/ionization (MALDI) and fast atom bombardment (FAB) data. The most accurate results, with an average ppm error of 9.2 ± 6.6 ppm, were reported by Canada using ESI interfaced coupled to a high resolution mass spectrometer. Figure 1 illustrates the high resolution ESI-MS data acquired for Peptide E by Canada. Using a resolution of 6000 (10% valley definition) it was possible to completely resolve the isotopic cluster for the $(M+6H)^{6+}$ ion enabling calculation of the monoisotopic molecular weight from the ion containing only ^{12}C. Even taking into account that a variety of instrumentation was used to determine molecular weights (Table II), the magnitude of the ppm errors between the theoretical and reported molecular weights was larger than expected. If molecular weight is to be used to reduce the number of possible hits during peptide database searching, then it would be advantageous to determine the molecular weights of unknown peptides with as high a degree of accuracy as possible.

Determination of Amino Acid Sequence

Peptide A: FIPK

Sequence data was submitted for Peptide A by Canada, the Netherlands, the United Kingdom, the United States #1 and the United States #2 laboratories indicating the sequence was FIPK. The laboratories from Canada, the United Kingdom and the United States #2 noted that under low energy collisionally activated dissociation (CAD) conditions they could not distinguish between the isomers of leucine and isoleucine (for clarity, the correct isomer has been assigned throughout the text by the authors). The Netherlands and United States #1 laboratories indicated that they could distinguish between isoleucine and leucine. The Netherlands assigned isoleucine by recording the product ion spectrum of m/z 86 (the immonium fragment of these amino acids) of the intact peptide (see discussion of Peptide D). The United States #1 laboratory assigned isoleucine based on spectral data from FAB-MS/MS under high energy CAD conditions and from data acquired with their automated peptide sequencer. Figure 2 illustrates the CAD/MS and MS/MS spectra obtained for Peptide A by Canada. Y_n-series ions were observed in the CAD/MS spectrum at m/z 357.2449 (y_3), m/z 244.1583 (y_2) and m/z 147.1046 (y_1) which allowed

assignment of the amino acid sequence. The protonated molecular ion was selected as the precursor ion and the product ion spectra acquired under MS/MS conditions (CAD cell 100V, $1.1x10^{-4}$ Torr Argon) contained y_n-, b_n- and a_n- series ions which were used to confirm the amino acid sequence. Most laboratories indicated that the C-terminal of Peptide A could be either lysine or glutamine. Accurate mass measurement under high resolution conditions was used by Canada to assign lysine as the C-terminal. The United Kingdom assumed lysine based on the strong doubly charged ion in the ES spectrum. The United States #1 laboratory assigned lysine based on data from their automated peptide sequencer. The United States #2 laboratory assigned lysine based on the observed increase in molecular weight (by 84 mass units) following acetylation with acetic anhydride.

Peptide B: AGKDYDKIEE

The Netherlands and the United States #1 laboratories submitted data for this peptide that indicated the complete sequence was AGKDYDKIEE. The United States #2 laboratory submitted data that indicated that the amino acid sequence was QKDYDKIEE (where Q has the same nominal mass of A and G). Partial sequence data was submitted for this peptide by Canada, which indicated that if the observed ions were due to b_n-series ions, the peptide contained the partial sequence IEE at the C-terminal. The United Kingdom submitted 2 possible sequences: QKFMDKIEE or KKFMDQIEE for Peptide B. Figure 3 illustrates the product spectrum for the $(M+H)^+$ precursor ion of Peptide B acquired by the United States #2 laboratory during ESI-MS/MS. It is the average of 24 scans and was acquired on a Finnigan-MAT TSQ-7000 with a cell voltage of 40V and an argon pressure of 3.5 mbar. The spectrum contains b_9 through b_2 series ions from which most of the amino acid sequence could be assigned.

Peptide C: ATKKEVPLGVAADANKLG

As molecular weight increases, the ability to produce and observe product ions decreases with triple quadrupole instruments due to the lower energies associated with the CAD process and reduced higher mass ion transmission. The typical solution is enzymatic or chemical digestion of the peptide to produce

lower mass fragments that may be sequenced. Only the United States #1 laboratory submitted mass spectrometric data that allowed complete amino acid sequencing of intact Peptide C. The Netherlands submitted data for tryptic fragments of this peptide that indicated the complete sequence was ATKKEVPLGVAADANKLG. Following tryptic digestion, partial sequence data was submitted for this peptide by Canada, which indicated that the peptide contained the partial internal sequence, PLGVA. The United States #2 laboratory following tryptic and chymotryptic digestions, submitted the amino acid sequence: ATKKEVPLRASVANQLG. The United Kingdom following tryptic digestion, submitted data that indicated that the peptide contained the partial sequence, PMPLGVAA. Figure 4 illustrates the FAB-MS/MS spectrum acquired by the United States #1 laboratory for Peptide C. B_n-series ions dominate the spectrum and although the complete series is not present, ions from b_{18} down to b_1 were observed. Under low energy collisional conditions, y_n-, b_n- and related series ions are typically observed. Under the high energy collisional conditions used by the United States #1 laboratory (2kV) in addition to these ions, an additional series of ions related to the loss of amino acid functional groups were also observed. This additional fragmentation while providing structural information, makes the product ion spectra increasingly more difficult to interpret.

Peptide D: RFEMFRELNEALELKDAQAGKE

No laboratories were able to provide sequence information on the intact peptide. The Netherlands and the United States # 1 laboratories submitted data for tryptic fragments for this peptide that indicated the complete sequence was RFEMFRELNEALELKDAQAGKE. The United Kingdom submitted two possible partial sequences for a tryptic fragment of peptide D, ELNEALELK or ELNEALELE-NH_2. It has generally been accepted that under low energy collisional conditions, that it is not possible to distinguish between the two isomers leucine and isoleucine. In this training exercise, the Netherlands laboratory reported a method by which it is possible to distinguish between these two amino acids. They recorded the ESI product ion spectrum of the immonium ion (m/z 86) of the intact peptide. The product spectrum was then compared to reference spectrum of both amino acids. Figure 5 illustrates the ESI spectra obtained for Peptide D under three cone voltages showing the relative intensity of the immonium ion increasing with cone voltage. Figure 6 illustrates the product ion spectra acquired for Peptide D, a

standard of leucine and a standard of isoleucine. If the peptide or peptide fragment (e.g. tryptic fragment) contains only leucine or isoleucine this approach offers a powerful means of correctly assigning the presence of these two amino acids.

Figure 7 illustrates the FAB-MS/MS spectrum obtained by the United States #1 laboratory for the GLU-C fragment MFRE of Peptide D. The spectrum contains the b_4, b_3, a_2 and a_1 series ion from which it is possible to assign the sequence for this fragment. Confirmation of this sequence is supplied by the y_3, y_2 and y_1 ions. Figure 8 illustrates the FAB-MS/MS spectrum of the chymotryptic fragment RELNEAL submitted by the United States #1 laboratory. The complete sequence can almost be assigned from the presence of the a_7 to a_2 series ions and was confirmed by data from their automated peptide sequencer. Both product spectra were obtained under high energy collisional conditions and as was the case for Peptide C exhibited additional series ions.

Peptide E: VEHYDNIEQKIDDIDHEIADLQAKITRLVQQHPRIDE

The United States #1 laboratory submitted data that indicated the complete sequence for Peptide E. This data was acquired with an automated peptide sequencer and was supported by mass spectrometric analysis of some proteolytic (trypsin, chymotrypsin and GLU-C) fragments. The Netherlands reported observing the presence of five tryptic fragments and submitted the following partial sequences: VEHY(DVQ), EQK, ITR, IVQQHPR and IDE. The Netherlands reported that the complete sequence for Peptide E could not be elucidated in the timeframe of five working days allocated for the international training exercise. The United Kingdom indicated the presence of three tryptic fragments with the partial sequence from one fragment: ALSHQQVP, but also noted that this sequence was not consistent with the expected cleavages associated with the use of trypsin.

Canada reported five tryptic fragments for Peptide E with monoisotopic molecular weights of 388.23, 375.16, 876.48, 1273.56 and 1594.78 Da (tryptic map). Figure 9 illustrates the total-ion-current chromatogram (1400 to 375 Da) and reconstructed-ion-current chromatograms, obtained by the Canadian laboratory, for the five tryptic fragments obtained during LC-ESI-MS analysis of Peptide E. The sum of

the masses of these five tryptic fragments was within 1 ppm of the monoisotopic molecular weight determined for the unknown peptide. Partial sequence data was submitted for three of the tryptic fragments for Peptide E by Canada, indicating that the peptide contained the partial sequences ITR, LVQQ and IDE (C-terminal).

CONCLUSIONS

The revision of the Terms of Reference for the SICA group to include the entire CB agent spectrum and the current lack of analytical methods for the identification of mid-spectrum agents was the impetus behind this SICA international training exercise. Seven laboratories from six countries participated in the training exercise whose objective was to evaluate the basic capabilities of the laboratories to determine molecular weights and amino acid sequences for five unknown peptides.

The determination of an unknown's molecular weight is considered to be the initial step in the identification of a mid-spectrum agent. All the participating laboratories demonstrated the basic ability to determine either monoisotopic or average molecular weights for the unknown peptide samples. The most accurate results, with an average ppm error of 9.2 ± 6.6 ppm, were reported by Canada using ESI interfaced coupled to a high resolution mass spectrometer. Even taking into account that a variety of instrumentation was used to determine molecular weights, the magnitude of the ppm errors between the theoretical and reported molecular weights was larger than expected. If molecular weight is to be used to reduce the number of possible hits during peptide database searching, then it would be advantageous to determine the molecular weights of unknown peptides with as high a degree of accuracy as possible.

The second objective of the training exercise was to determine partial or complete amino acid sequences for the unknown peptides. For a number of the participating laboratories, this was their first experience in analyzing peptides. They indicated that they had limited themselves to determining molecular weights as they did not feel that they currently possessed the expertise to provide amino acid sequence data. The other laboratories provided either partial or complete amino acid sequences for the unknown peptides. The only laboratory to provide the correct amino acid sequence for all five peptides

was the United States #1 laboratory. They were able to sequence the unknown peptides by using both automated peptide sequencer and mass spectrometric data. The laboratory from the Netherlands using only mass spectrometry was able to completely sequence four unknown peptides and provided partial sequence data for the fifth peptide. The Netherlands also reported a procedure for distinguishing between the isomers of isoleucine and leucine which had not been thought possible under low energy collisonally activated dissociation conditions.

From this training exercise, it is clear that the use of an automated peptide sequencer (e.g. Edmans degradation) similar to that used by the United States #1 laboratory may prove invaluable. Sequencers are not without their limitations, as the N-terminal of a peptide maybe modified thereby blocking the terminal from enzymatic cleavage. The general consensus is that between 25 and 50% of natural occurring peptides are blocked from sequencing using Edmans degradation.

The use of enzymes such as trypsin, chymotrypsin, carboxy peptidases (which provide sequencing from the C-terminal) and Glu-C provide another means to access amino acid sequence information. The enzymatic fragments being of lower mass than the intact peptide can in some cases be sequenced directly by mass spectrometry. The molecular weights of the enzymatic fragments (mass map) together with the molecular weight of the intact peptide can be used to search peptide databases. Chemical derivatization may be used to access additional information such as the number of acid sites in the peptide.

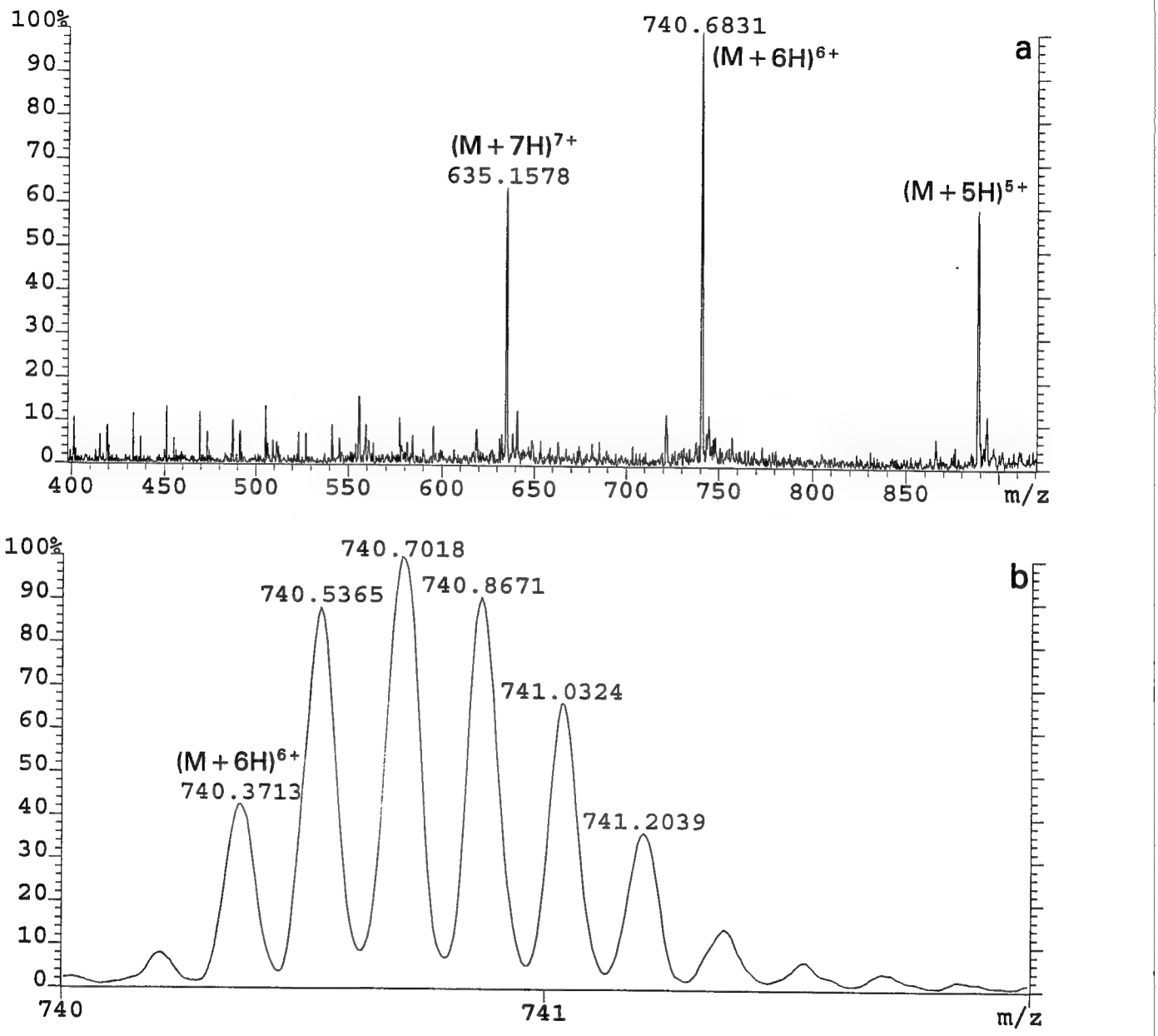
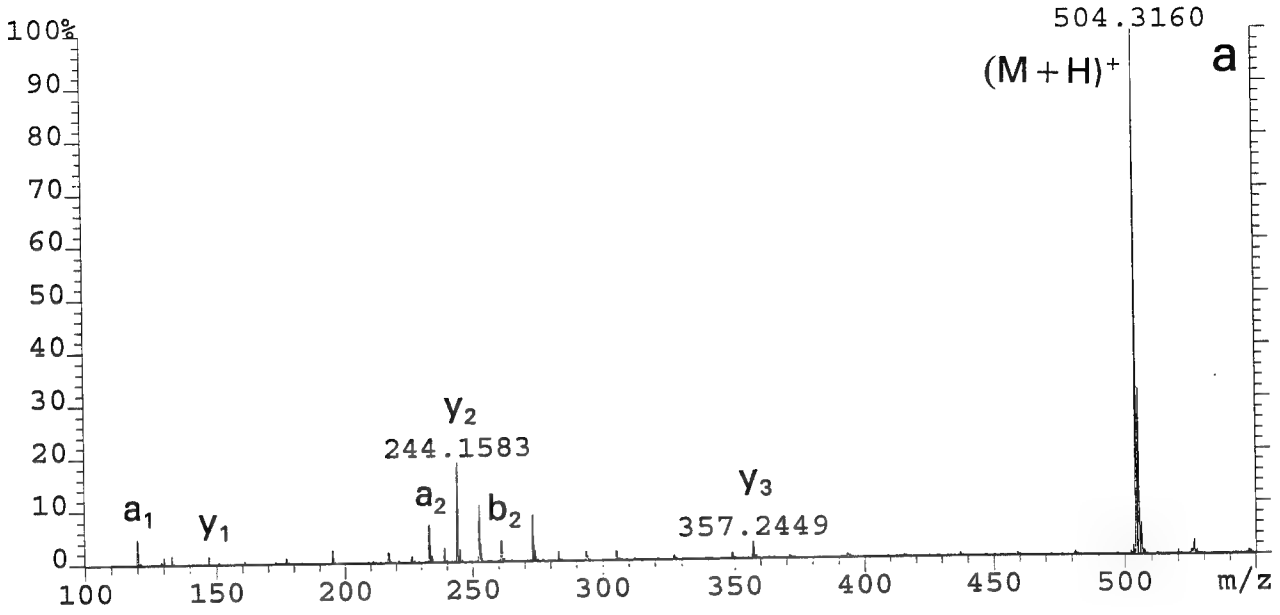


Figure 1: a) High resolution ESI-MS data acquired for Peptide E by Canadian laboratory during magnetic sector scanning over a wide mass range [925 to 400 Da]. b) $(M+6H)^{6+}$ isotopic cluster acquired during magnetic sector scanning over a narrow mass range [790 to 660 Da]. MS Conditions: Sampling cone voltage, 25 volts; Resolution, 6000 (10% valley). Calculated monoisotopic molecular weight: 4436.172±0.034 Da (n=4).



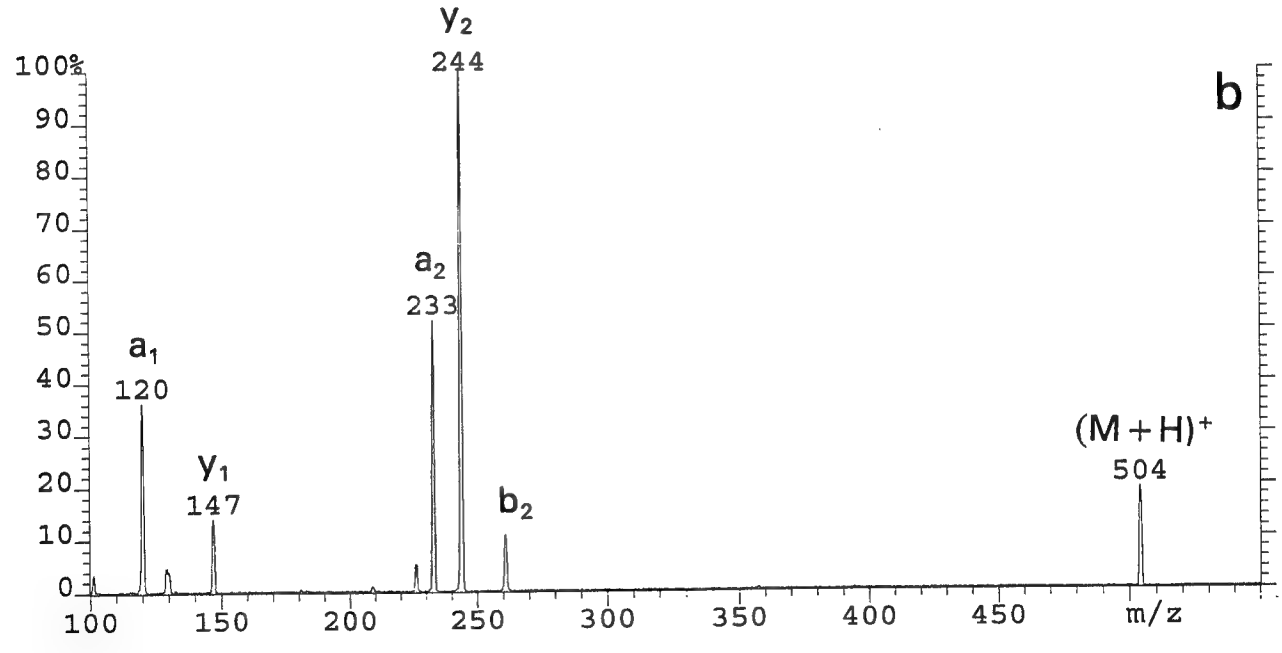


Figure 2: a) CAD/MS (75 volts) and b) ESI-MS/MS (Quadrupole CAD cell; 100 volts, 1.1×10^{-4} Torr Argon) spectra obtained by Canadian laboratory during analysis of Peptide A.

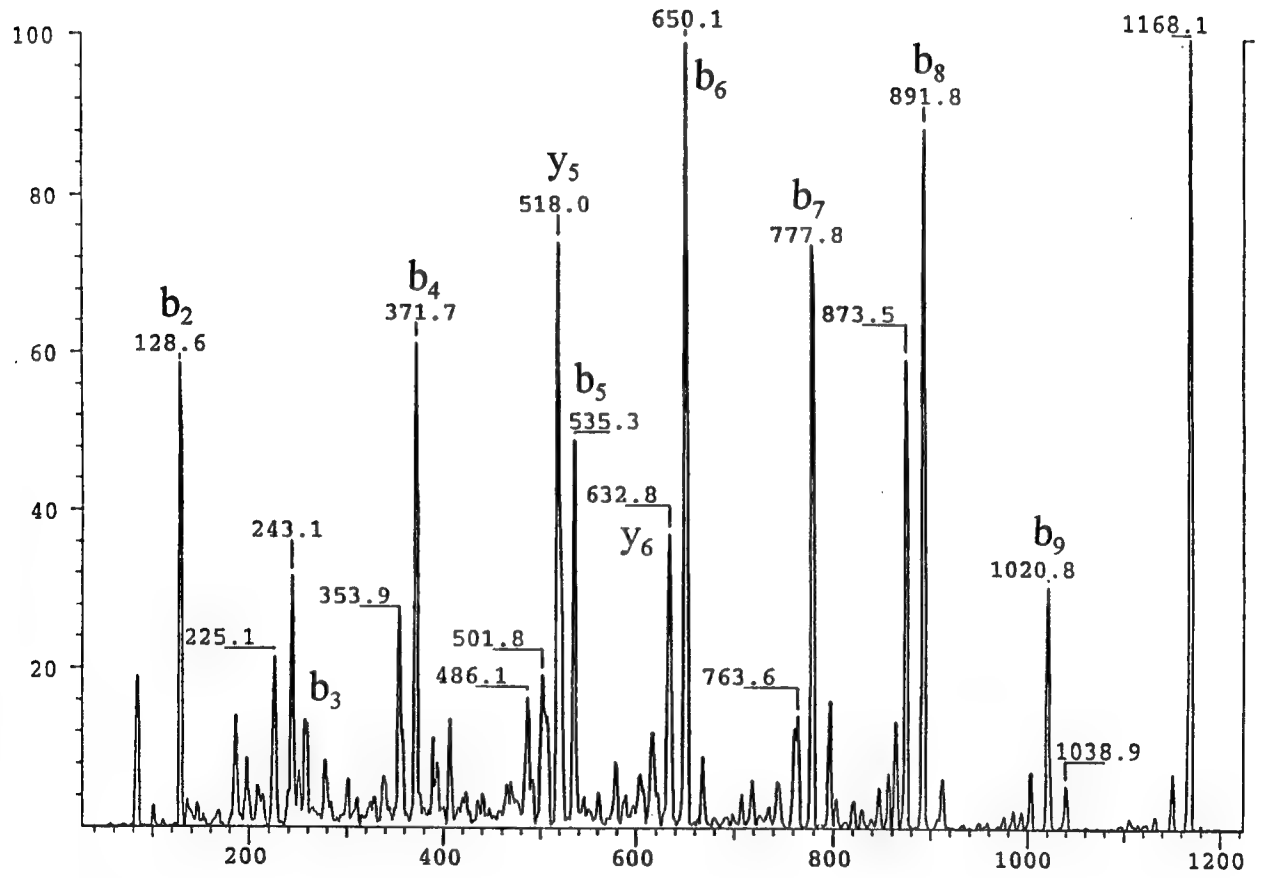


Figure 3: Product ion spectrum obtained by United States #2 laboratory during ESI-MS/MS (CAD Cell; 40V, 3.5mTorr Argon) analysis of Peptide B.

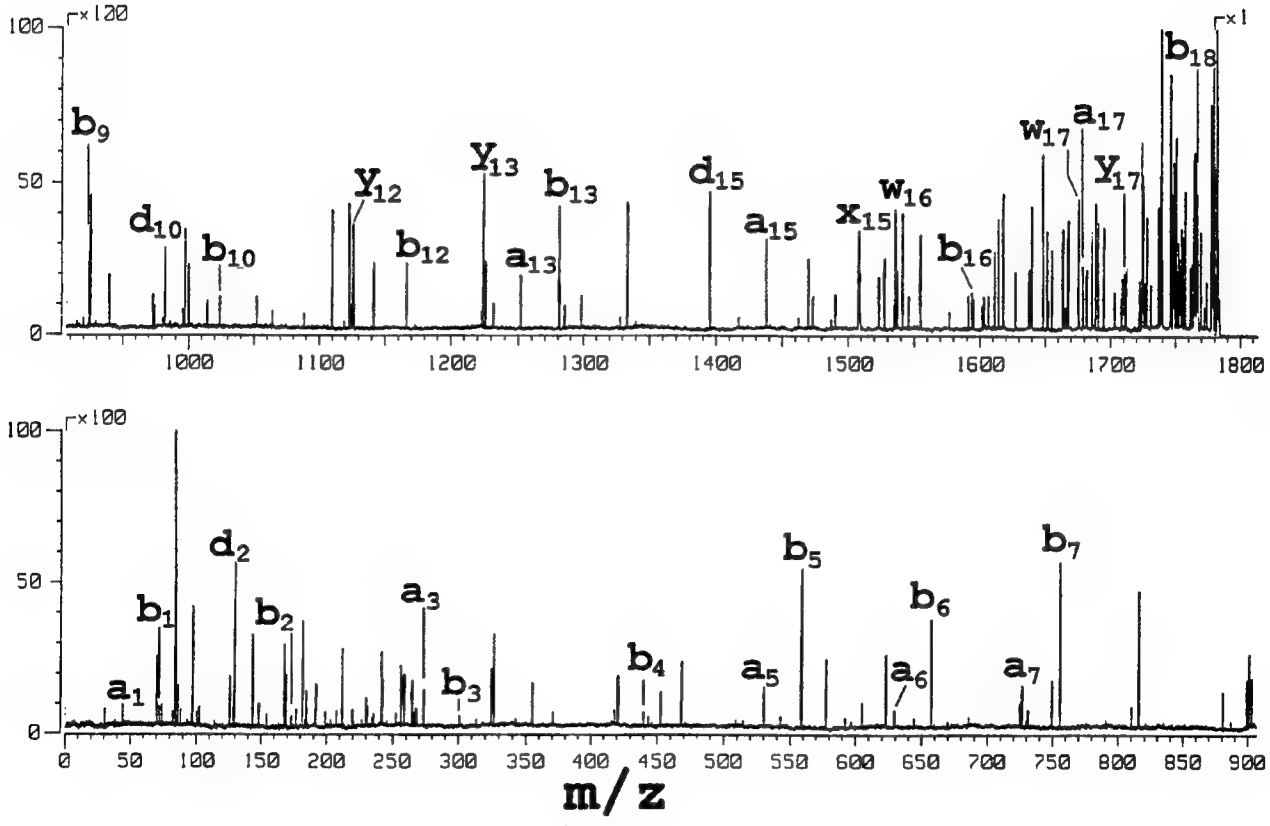


Figure 4: Product ion spectrum obtained by United States #1 laboratory during FAB-MS/MS (CAD Cell; 2kV, 1.5 mTorr Xenon) analysis of Peptide C.

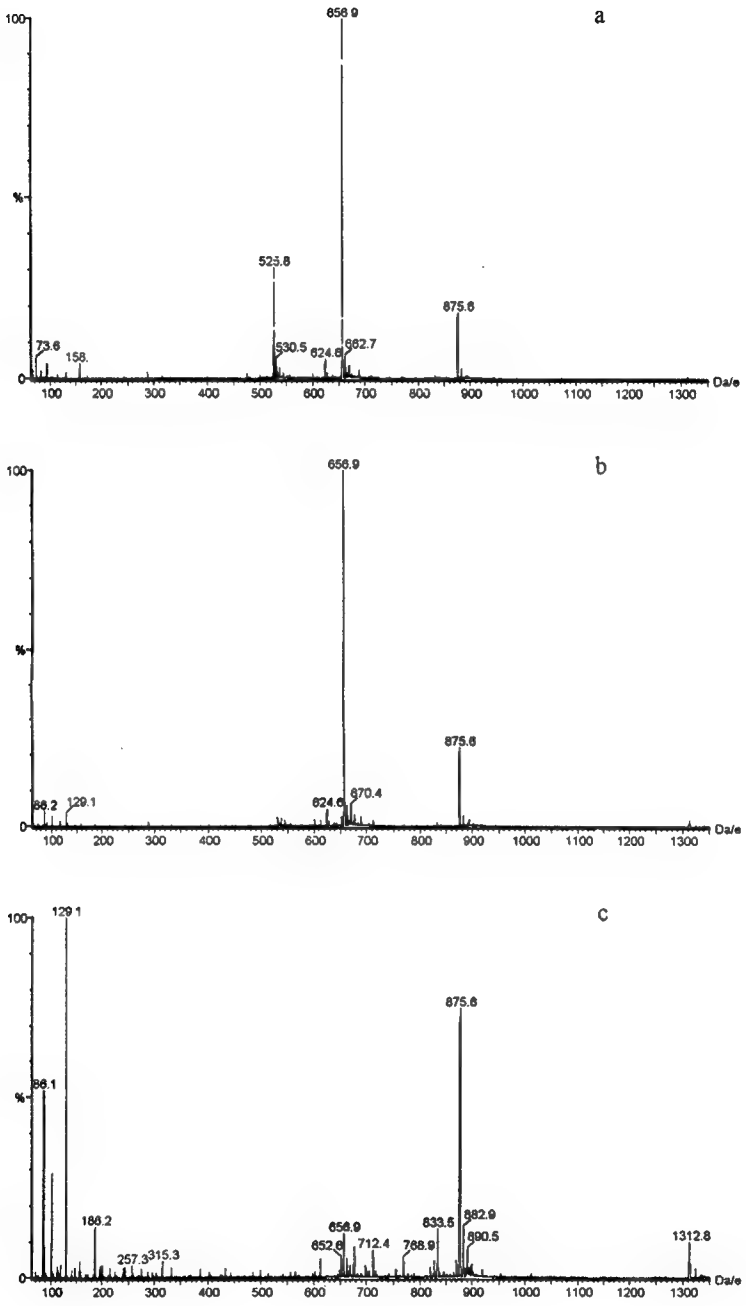


Figure 5: ESI-MS data obtained by Netherlands laboratory for Peptide D at three cone voltages; a) 35V, b) 50V and c) 75V.

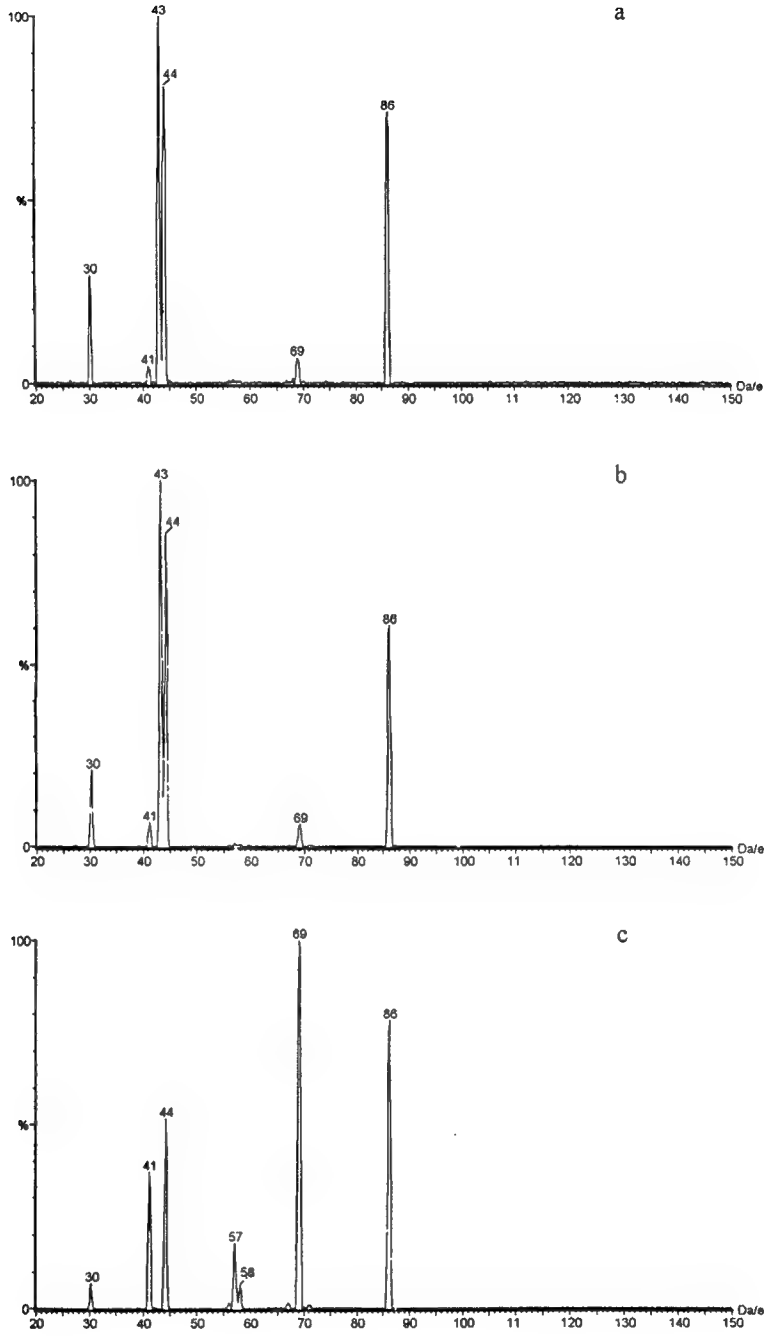


Figure 6: Product ion spectra obtained by Netherlands laboratory for immoniun ion (m/z 86) during ESI-MS/MS (CAD cell 12V, 5 x 10^{-3} mbar Argon) analysis of ; a) Peptide D, b) leucine and c) isoleucine.

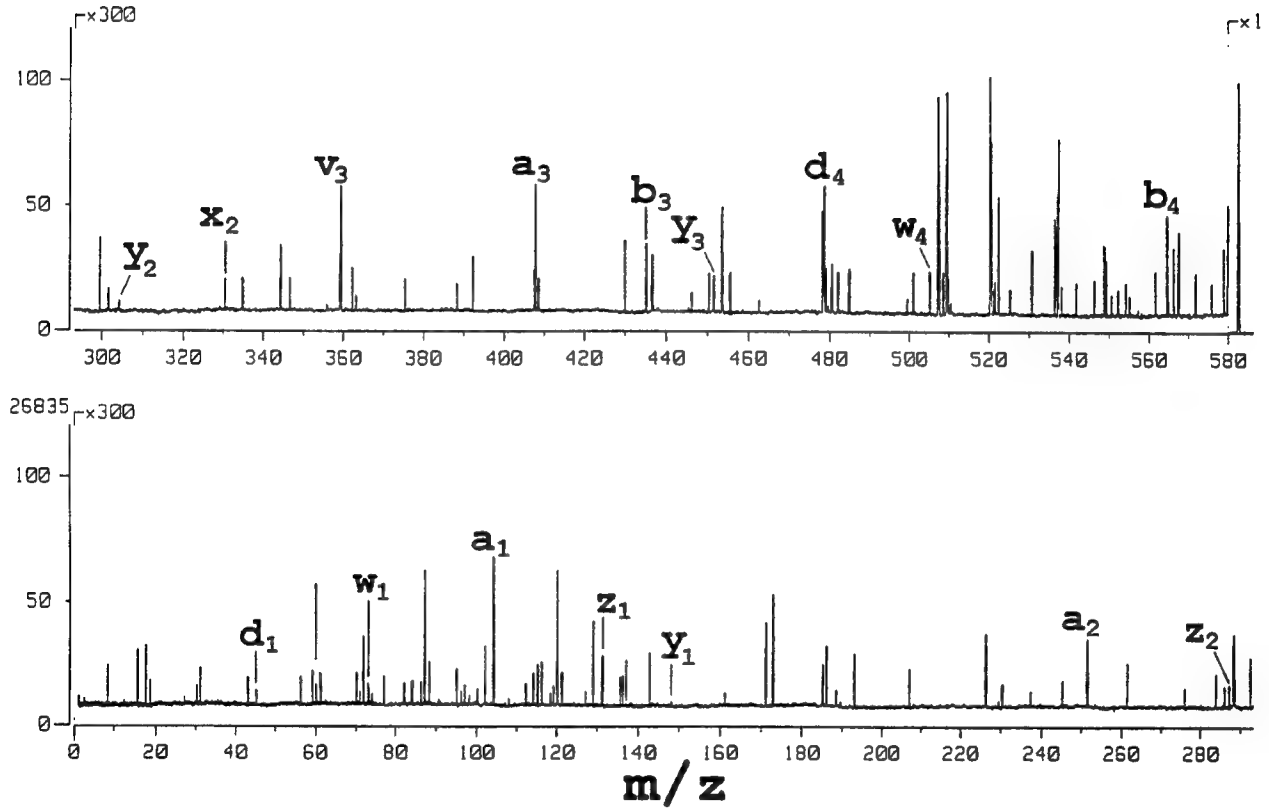


Figure 7: Product ion spectrum obtained by United States #1 laboratory during FAB-MS/MS (CAD Cell; 2kV, 1.5 mTorr Xenon) analysis of Glu-C fragment MFRE of Peptide D.

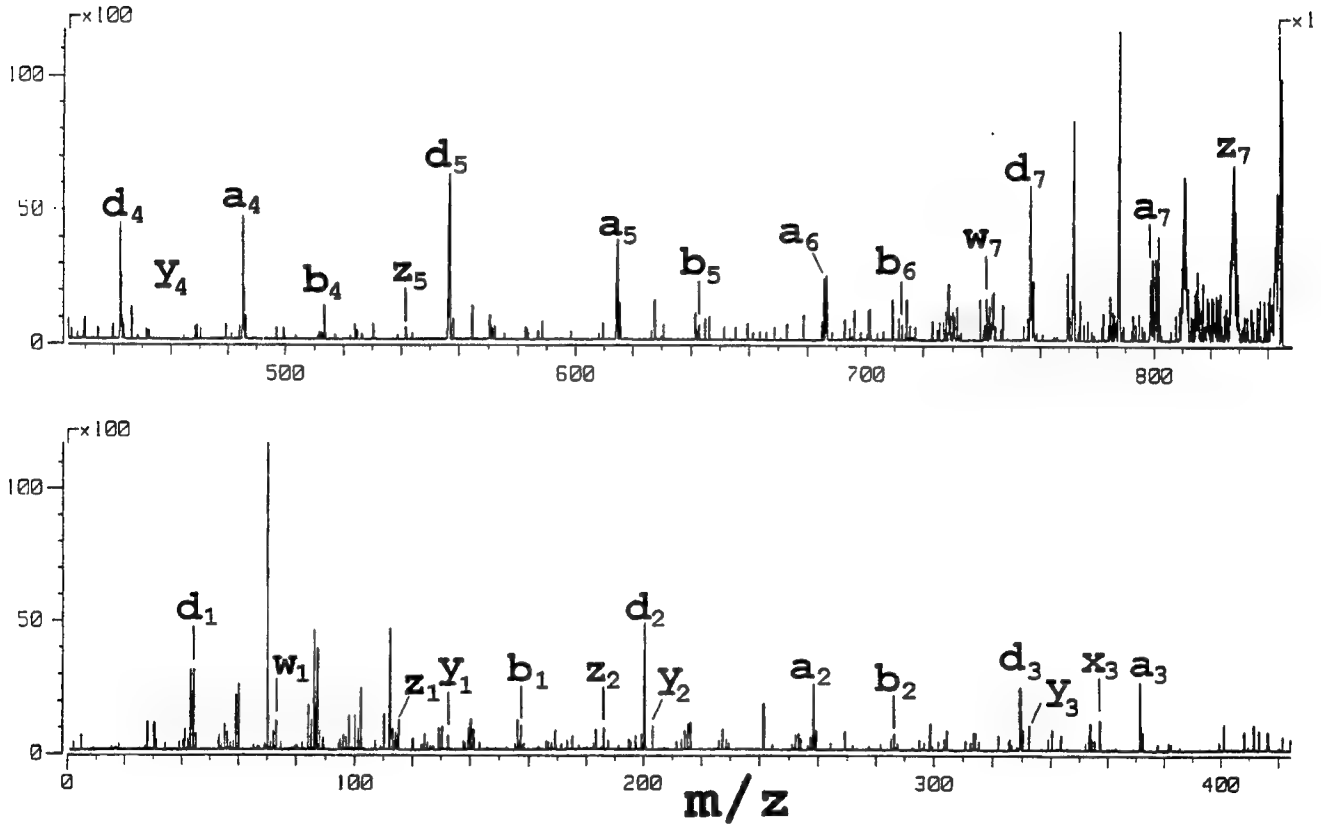


Figure 8: Product ion spectrum obtained by United States #1 laboratory during FAB-MS/MS (CAD Cell; 2kV, 1.5 mTorr Xenon) analysis of chymotryptic fragment RELNEAL of Peptide D.

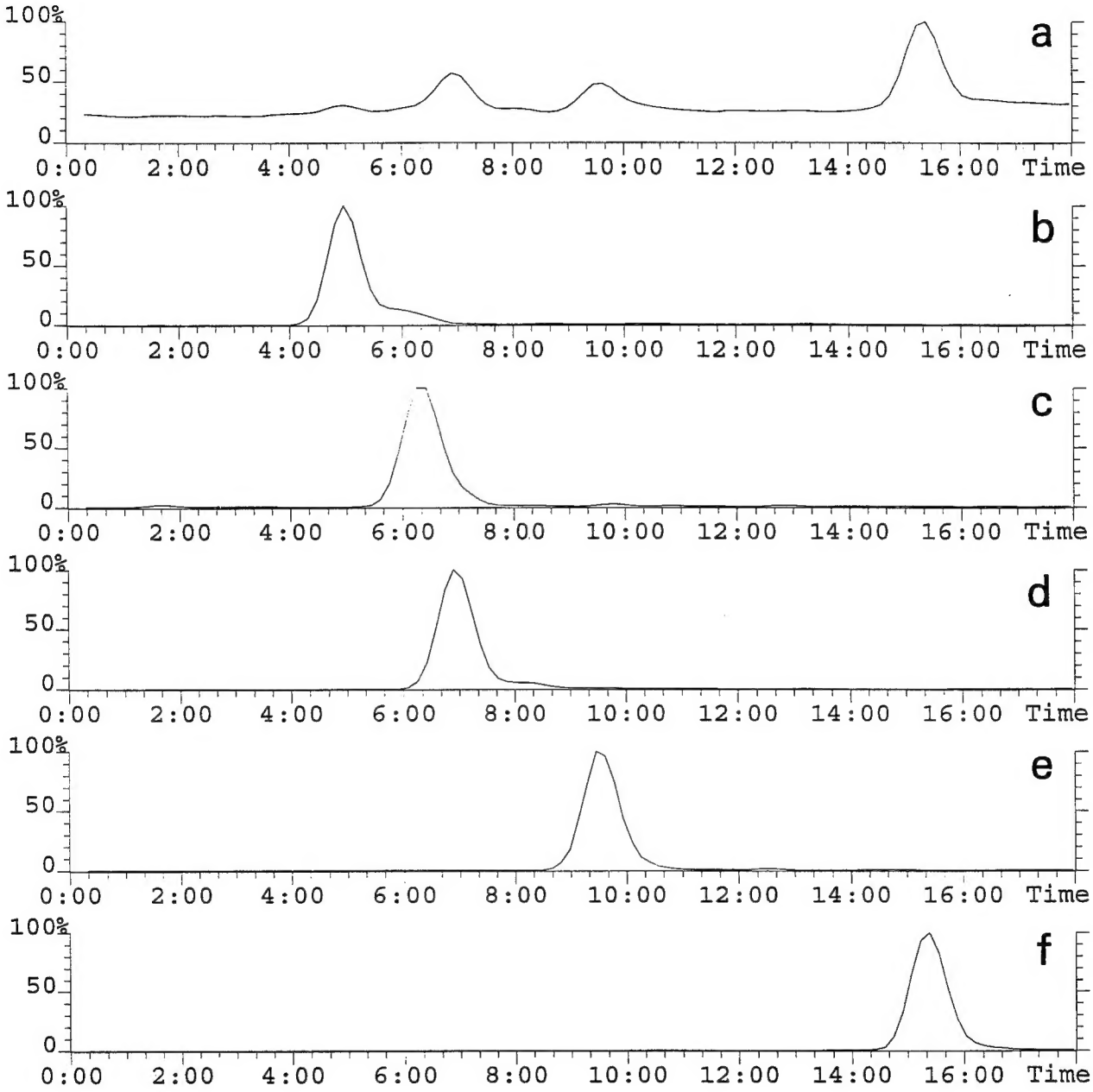


Figure 9: a) Total-Ion-Current (1400 to 375 Da) chromatogram for tryptic digest of Peptide E obtained by Canadian laboratory during LC-ESI-MS analysis. Reconstructed-Ion-Current chromatograms for b) m/z 389.2, c) m/z 376.2, d) m/z 877.5, e) m/z 637.8 and f) m/z 798.4 (Masses of tryptic fragments of Peptide E).

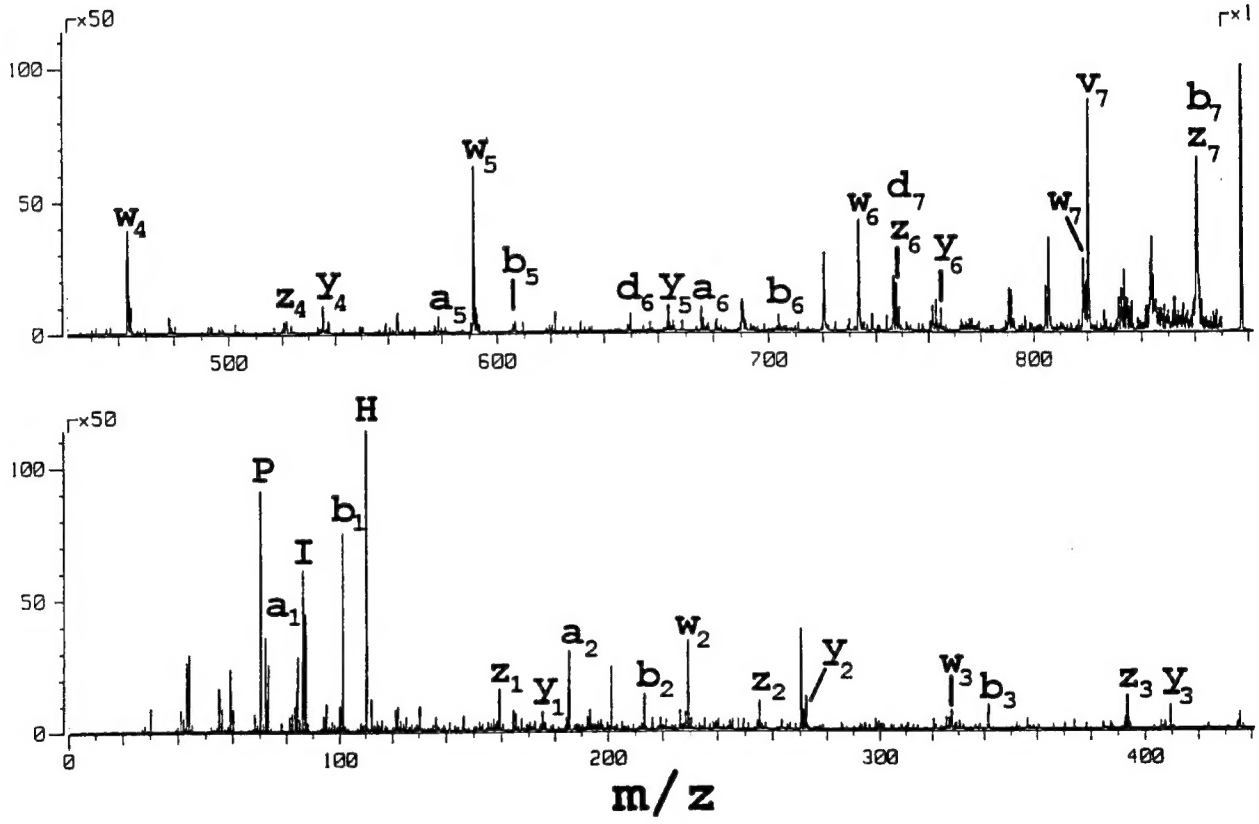


Figure 10: Product ion spectrum obtained by United States #1 laboratory during FAB-MS/MS (CAD Cell; 2kV, 1.5 mTorr Xenon) analysis of tryptic fragment LVQQHPR of Peptide E.

DOCUMENT CONTROL DATA

(Security classification of title, body of abstract and indexing annotation must be entered when the overall document is classified)

1. ORIGINATOR (the name and address of the organization preparing the document. Organizations for whom the document was prepared, e.g. Establishment sponsoring a contractor's report, or tasking agency, are entered in section 8.)

Defence Research Establishment Suffield
Box 4000
Medicine Hat, AB T1A 8K6

2. SECURITY CLASSIFICATION (overall security classification of the document. including special warning terms if applicable)

Unclassified

3. TITLE (the complete document title as indicated on the title page. Its classification should be indicated by the appropriate abbreviation (S,C,R or U) in parentheses after the title.)

NATO Panel VII/SICA Training Exercise on the Identification of Peptides

4. AUTHORS (Last name, first name, middle initial. If military, show rank, e.g. Doe, Maj. John E.)

Hancock, James R. and D'Agostino, Paul A.

5. DATE OF PUBLICATION (month and year of publication of document)	6a. NO. OF PAGES (total containing information. Include Annexes, Appendices, etc.)	6b. NO. OF REFS (total cited in document)
May 1996	21	0

6. DESCRIPTIVE NOTES (the category of the document, e.g. technical report, technical note or memorandum. If appropriate, enter the type of report, e.g. interim, progress, summary, annual or final. Give the inclusive dates when a specific reporting period is covered.)

Suffield Memorandum

8. SPONSORING ACTIVITY (the name of the department project office or laboratory sponsoring the research and development. Include the address.)

9a. PROJECT OR GRANT NO. (if appropriate, the applicable research and development project or grant number under which the document was written. Please specify whether project or grant)	9b. CONTRACT NO. (if appropriate, the applicable number under which the document was written)
351SQ	

10a. ORIGINATOR'S DOCUMENT NUMBER (the official document number by which the document is identified by the originating activity. This number must be unique to this document.)	10b. OTHER DOCUMENT NOS. (Any other numbers which may be assigned this document either by the originator or by the sponsor)
SM 1481	

11. DOCUMENT AVAILABILITY (any limitations on further dissemination of the document, other than those imposed by security classification)

- (X) Unlimited distribution
- ( ) Distribution limited to defence departments and defence contractors; further distribution only as approved
- ( ) Distribution limited to defence departments and Canadian defence contractors; further distribution only as approved
- ( ) Distribution limited to government departments and agencies; further distribution only as approved
- ( ) Distribution limited to defence departments; further distribution only as approved
- ( ) Other (please specify):

12. DOCUMENT ANNOUNCEMENT (any limitation to the bibliographic announcement of this document. This will normally correspond to the Document Availabilty (11). However, where further distribution (beyond the audience specified in 11) is possible, a wider announcement audience may be selected.)

13. ABSTRACT (a brief and factual summary of the document. It may also appear elsewhere in the body of the document itself. It is highly desirable that the abstract of classified documents be unclassified. Each paragraph of the abstract shall begin with an indication of the security classification of the information in the paragraph (unless the document itself is unclassified) represented as (S), (C), (R), or (U). It is not necessary to include here abstracts in both offical languages unless the text is bilingual).

The revision of the Terms of Reference for NATO's Sampling and Identification of Chemical/Biological Agents (SICA) subgroup to include the entire CB agent spectrum and the current lack of analytical methods for mid-spectrum agents prompted SICA to hold an international training exercise focussing on the identification of these agents. From the 11 NATO countries represented on the SICA subpanel, laboratories from Canada (host nation), Denmark, Netherlands, Norway, United Kingdom and United States (2 laboratories) agreed to participate in the international training exercise. The objective of the exercise was to evaluate the basic capabilities of the laboratories to determine molecular weights and primary amino acid sequences for five unknown peptides with molecular weights in the same mass range as many mid-spectrum agents.

The determination of an unknown's molecular weight is considered to be one of the initial steps in the identification of a mid-spectrum agent. All the participating laboratories demonstrated the basic ability to determine either monoisotopic or average molecular weights for the unknown peptide samples. The most accurate results, with an average error of 9.2 ppm, in this part of the training exercise were reported by Canada, using electrospray ionization interfaced to a high resolution mass spectrometer.

The second objective of the training exercise was to determine partial or complete amino acid sequences for the unknown peptides. The only laboratory to provide the correct amino acid sequence for all five peptides was the United States #1 laboratory. They were able to sequence the unknown peptides by using a combination of an automated peptide sequencer and sequence data provided by mass spectrometry. The laboratory from the Netherlands using only mass spectrometry was able to completely sequence four of the unknown peptides and provided partial sequence data for the fifth peptide.

14. KEYWORDS, DESCRIPTORS or IDENTIFIERS (technically meaningful terms or short phrases that characterize a document and could be helpful in cataloguing the document. They should be selected so that no security classification is required. Identifiers, such as equipment model designation, trade name, military project code name, geographic location may also be included. If possible keywords should be selected from a published thesaurus. e.g. Thesaurus of Engineering and Scientific Terms (TEST) and that thesaurus-identified. If it is not possible to select indexing terms which are Unclassified, the classification of each should be indicated as with the title.)

NATO

Sampling and Identification of Chemical/Biological Agents

Peptides

Mass Spectrometry

Liquid Chromatography